

**ASSESSMENT ANALYTICS OF DIFFERENTIAL ITEM  
FUNCTIONING IN BASIC SCIENCE AND  
TECHNOLOGY MULTIPLE-CHOICE ITEMS OF THE  
2022 BASIC EDUCATION CERTIFICATE  
EXAMINATION IN ZONE C, BENUE STATE, NIGERIA**

**<sup>1</sup>OPITA MATTHEW JOHN**

**Department of Integrated Science Education, School of  
Sciences, Federal College of Education, Odugbo, Benue State.  
<sup>1</sup>opitamatthewjohn@gmail.com 08068681392**

**&**

**<sup>2</sup>UDEANI GLOARIA CHINYERE**

**Department of Mathematics Education, School of Sciences,  
Federal College of Education, Odugbo, Benue State.  
<sup>2</sup>gloriaudeani47@gmail.com**

**&**

**<sup>3</sup>IKWUJE ANDREW IKWUJE**

**Department of Integrated Science Education, School of  
Sciences, Federal College of Education, Odugbo, Benue State.  
<sup>3</sup>dbestikwuje@gmail.com**

**Abstract**

*The study investigated the Differential Item Functioning (DIF) of multiple-choice items in the Basic Science and Technology Basic Education Certificate Examination (BECE) in Zone C of Benue State, Nigeria. The study was motivated by persistent performance disparities across demographic groups, which raised concerns about the fairness and validity of the assessment. Although several factors may influence assessment fairness, this*

*study focused specifically on the assessment instrument. Three research questions guided the study. An ex-post facto research design was adopted. The population comprised 16,062 candidates, from which a sample of 3,142 students was purposively selected. The instrument for data collection consisted of students' response sheets containing answers to the 60 multiple-choice items of the 2022 BECE Basic Science and Technology examination. The validity and reliability of the instrument were ensured through the examination body's established quality assurance procedures; therefore, no additional validation was conducted by the researchers. Data were analysed using the Cochran–Mantel–Haenszel (CMH) method in JMetrik. The findings revealed that gender-based DIF was modest, with 9 items (15%) exhibiting significant bias. Location-based DIF was the most pronounced, with 42 items (70%) showing significant bias, while school-type DIF was also substantial, with 33 items (55%) exhibiting significant bias. The study concluded that although gender fairness was generally adequate, substantial location- and school type-related bias threatens the validity and fairness of the examination. It was therefore recommended that the Benue State Examinations Board institutionalize routine DIF analysis as part of its post-examination review process to identify and eliminate biased items, while the Federal and Benue State Ministries of Education should address structural educational inequalities across school locations to promote equitable assessment outcomes.*

**Keywords:** Differential Item Functioning, BECE, Test Fairness, Basic Science and Technology

## **Introduction**

Basic Science and Technology is an evolution from Integrated Science, serving as the foundational science course for upper basic level students (Upper Basic I–III) in Nigeria. Regardless of individual interests, all students are required to study it, as it

prepares them for advanced disciplines Biology, Chemistry, and Physics at the senior secondary level (Oguguo & Lotobi, 2019). The subject introduces scientific concepts in a simplified, coherent manner aligned with students' cognitive development, emphasizing practical activities, observation, and problem-solving to connect scientific knowledge to everyday life. The Nigerian Educational Research and Development Council (NERDC, 2016) outlines the curriculum objectives for this subject as: developing students' interest in science and technology; acquiring basic knowledge and skills; applying scientific knowledge to everyday situations; building problem-solving and critical thinking skills; and fostering positive scientific attitudes. These objectives underpin preparation for further science studies and participation in a technology-driven society, and their attainment depends heavily on effective instruction and assessment.

Assessment is the systematic process of gathering and interpreting information about student learning to inform instructional decisions and evaluate educational progress (Adejoh & Obinne, 2015). In science education specifically, it is essential for tracking progress, identifying learning gaps, and aligning teaching with national standards (Okpala, Onocha & Oyedeki, 2016). Critically, assessment quality depends on the instruments used poorly designed tools can yield misleading results and undermine the validity of score interpretations. One widely used assessment instrument is the Basic Education Certificate Examination (BECE), which relies predominantly on objective tests — particularly multiple-choice questions (MCQs). MCQs are valued for their reliability, efficiency, and capacity to sample a broad range of content, making them well-suited for large-scale assessments (Popham, 2016; Olutola, 2020). Each item consists of a stem, a set of options, and a key (correct answer), allowing consistent evaluation of students' knowledge, comprehension, and application of concepts. However, the widespread use of MCQs in high-stakes examinations raises concerns about measurement

fairness, specifically whether items function equitably across different student subgroups. This concern is addressed through Differential Item Functioning (DIF) analysis — a psychometric approach that identifies items which perform differently for subgroups of equal ability. Scholars define DIF as a statistical phenomenon where test items show differential probabilities of correct response for individuals from distinct demographic groups despite comparable proficiency levels (Martinková & Drabinová, 2018; Osterlind & Everson, 2019). Chalmers (2018) frames it as a threat to test fairness when item characteristics systematically advantage or disadvantage particular subgroups, while Magis, Béland, and Raïche (2020) position DIF analysis as an essential quality-control mechanism for protecting the validity of score interpretations across diverse populations. The relevance of DIF analysis is underscored by observable performance disparities in Benue State's BECE results from 2020 to 2022: male students consistently outperformed female students by 12–15%, urban students scored approximately 18% higher than rural peers, and students from co-educational schools outperformed those from non-co-educational schools by about 10%. While these gaps may partly reflect genuine ability differences, they may also signal the presence of biased items that systematically favour certain groups. This study therefore examines the 2022 BECE Basic Science and Technology Achievement Test for DIF across demographic subgroups in Zone C of Benue State, providing empirical evidence on whether the test functions equitably for all students

### **Statement of the Problem**

Basic Education Certificate Examination (BECE) is expected to provide a fair, valid, and reliable measure of students' achievement in Basic Science and Technology. For this to hold, every test item should give students of equal ability an equal chance of answering correctly, regardless of gender, school location, or school type ensuring that scores reflect true academic ability rather than demographic or school-related advantages.

Evidence from Zone C of Benue State, however, suggests this standard may not be fully met. Preliminary analysis of BECE results from 2020 to 2022 revealed persistent performance disparities: male students consistently outperformed female students, urban school students outperformed rural counterparts, and students from co-educational schools scored higher than those from non-co-educational schools. While these gaps may partly reflect differences in learning opportunities and academic preparation, they may also signal the presence of biased test items specifically, items exhibiting Differential Item Functioning (DIF). DIF occurs when students from different groups who possess the same level of ability have unequal probabilities of answering a test item correctly, systematically advantaging some groups while disadvantaging others. When DIF is present, test scores no longer represent students' true scientific knowledge and ability, undermining the validity of decisions made of those scores including placement, certification, and progression. Despite the importance of test fairness in large-scale assessment, limited empirical evidence exists on whether the 2022 BECE Basic Science and Technology Achievement Test functions equitably across these subgroups in Zone C of Benue State. This study therefore investigates whether the test's multiple-choice items exhibit DIF across gender (male and female), school location (urban and rural), and school type (co-educational and non-co-educational) among Upper Basic III students in the zone.

### **Objectives of the Study**

The study investigated Differential Item Functioning of Basic Science and Technology Multiple Choice Questions in the 2022 Basic Education Certificate Examination in Zone C of Benue State, Nigeria. The specific objectives of the study include, to:

- a. identify the items in the BECE 2022 Basic Science and Technology Multiple Choice Test that function differentially based on gender in Zone C of Benue State, Nigeria.

- b. identify the items in the BECE 2022 Basic Science and Technology Multiple Choice Test that function differentially based on location in Zone C of Benue State, Nigeria.
- c. ascertain items in the BECE 2022 Basic Science and Technology Multiple Choice Test that function differentially by school types (co-educational and non-co-educational) in Zone C of Benue State, Nigeria.

### **Research Questions**

The study sought answers to the following research questions:

1. How many items in the BECE 2022 Basic Science and Technology Multiple Choice Test function differentially based on gender in Zone C of Benue State Nigeria?
2. How many items in the BECE 2022 Basic Science and Technology Multiple Choice Test function differentially based on location in Zone C of Benue State Nigeria?
3. How many items in the BECE 2022 Basic Science and Technology Multiple Choice Test function differentially based on school types (co-educational and non-co-educational) in Zone C of Benue State Nigeria?

### **Methodology**

The study adopted an ex-post facto research design, which was appropriate given that the data existed before the commencement of the research. The demographic variables of interest gender, school location, and school type had already occurred naturally and were not subject to manipulation or intervention (EmAikwu, 2019). The study was conducted in Zone C of Benue State, Nigeria. The target population comprised 16,062 Upper Basic III (JSS 3) students who sat the 2022 BECE Basic Science and Technology multiple-choice test in Zone C of Benue State, drawn from both co-educational and non-co-educational schools across urban and rural locations (BSEB, 2022). A sample of 3,142 students was selected. Purposive sampling was employed to select participants with complete item response data across all 60

test items and verifiable demographic information on gender, school location, and school type. Students whose records were incomplete, illegible, or lacked the required demographic details were excluded. The final sample was subsequently disaggregated across the three demographic subgroups for analysis. The instrument for data collection was the BECE 2022 Basic Science and Technology Objective Answer Sheet, administered by the Benue State Examinations Board (BSEB). Each answer sheet captured students' responses to all 60 multiple-choice items alongside demographic information (gender, school name, and local government area). Responses were dichotomously coded: 1 for a correct response and 0 for an incorrect or blank response, in accordance with the official BECE 2022 Basic Science marking scheme. Since the test was a standardized official examination developed by the BSEB following established test construction procedures, no additional validation was conducted by the researcher. Reliability was similarly assured through the standard psychometric procedures applied by the examination body during test development. Data were analysed using the Cochran-Mantel-Haenszel (CMH) method, implemented in JMetrik software, to detect differential item functioning (DIF) across gender (male/female), school location (urban/rural), and school type (co-educational/non-co-educational). The CMH method is a widely used non-parametric approach for DIF detection that compares the item performance of focal and reference groups after conditioning on total test scores as a proxy for ability.

## **Results**

### **Research Question One**

How many items in the BECE 2022 Basic Science Multiple Choice Test function differentially based on gender in Zone C of Benue State, Nigeria?

To answer this research question, Cochran-Mantel-Haenszel DIF in jmetrik was used.

**Table 1. Presents the Cochran-Mantel-Haenszel DIF analysis results for gender groups (Males as reference group, Females as focal group)**

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta$ MH) | 95% C.I.       | DIF Class | Direction       |
|------|----------|---------|----------------------------|----------------|-----------|-----------------|
| qs1  | 14.69    | .00     | 1.98                       | (3.04, 0.91)   | B+        | Favours Males   |
| qs2  | 0.00     | .98     | 0.02                       | (1.54, -1.49)  | A         | No DIF          |
| qs3  | 2.66     | .10     | 0.85                       | (1.87, -0.16)  | A         | No DIF          |
| qs4  | 1.30     | .25     | 0.44                       | (1.19, -0.32)  | A         | No DIF          |
| qs5  | 1.16     | .28     | -0.56                      | (0.45, -1.56)  | A         | No DIF          |
| qs6  | 3.80     | .05     | 0.63                       | (1.26, -0.00)  | A         | No DIF          |
| qs7  | 0.17     | .68     | 0.17                       | (0.97, -0.64)  | A         | No DIF          |
| qs8  | 0.02     | .89     | 0.05                       | (0.75, -0.65)  | A         | No DIF          |
| qs9  | 0.30     | .58     | -0.18                      | (0.45, -0.81)  | A         | No DIF          |
| qs10 | 0.11     | .73     | 0.11                       | (0.72, -0.51)  | A         | No DIF          |
| qs11 | 1.04     | .31     | -0.33                      | (0.31, -0.97)  | A         | No DIF          |
| qs12 | 3.31     | .07     | 0.75                       | (1.56, -0.06)  | A         | No DIF          |
| qs13 | 3.50     | .06     | 0.68                       | (1.39, -0.03)  | A         | No DIF          |
| qs14 | 3.29     | .07     | 0.57                       | (1.21, -0.06)  | A         | No DIF          |
| qs15 | 1.31     | .25     | 0.42                       | (1.13, -0.29)  | A         | No DIF          |
| qs16 | 14.37    | .00     | 1.29                       | (1.97, 0.62)   | B+        | Favours Males   |
| qs17 | 9.28     | .00     | -1.27                      | (-0.45, -2.08) | B-        | Favours Females |
| qs18 | 0.91     | .34     | -0.32                      | (0.33, -0.96)  | A         | No DIF          |
| qs19 | 14.09    | .00     | -1.19                      | (-0.56, -1.81) | B-        | Favours Females |
| qs20 | 2.17     | .14     | 0.48                       | (1.12, -0.17)  | A         | No DIF          |
| qs21 | 3.87     | .05     | -0.68                      | (0.00, -1.36)  | A         | No DIF          |
| qs22 | 8.82     | .00     | 0.92                       | (1.53, 0.31)   | A         | No DIF          |
| qs23 | 4.00     | .05     | 0.70                       | (1.40, 0.00)   | A         | No DIF          |

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta MH$ ) | 95% C.I.       | DIF Class | Direction     |
|------|----------|---------|-----------------------------|----------------|-----------|---------------|
| qs24 | 25.51    | .00     | 1.81                        | (2.52, 1.09)   | C+        | Favours Males |
| qs25 | 0.24     | .62     | -0.20                       | (0.59, -0.99)  | A         | No DIF        |
| qs26 | 0.01     | .94     | 0.02                        | (0.67, -0.62)  | A         | No DIF        |
| qs27 | 0.25     | .61     | -0.16                       | (0.48, -0.81)  | A         | No DIF        |
| qs28 | 0.58     | .45     | -0.26                       | (0.42, -0.95)  | A         | No DIF        |
| qs29 | 4.79     | .03     | -0.71                       | (-0.07, -1.35) | A         | No DIF        |
| qs30 | 19.49    | .00     | 1.63                        | (2.37, 0.89)   | B+        | Favours Males |
| qs31 | 0.40     | .53     | 0.28                        | (1.16, -0.59)  | A         | No DIF        |
| qs32 | 1.31     | .25     | 0.37                        | (1.02, -0.27)  | A         | No DIF        |
| qs33 | 0.46     | .50     | -0.24                       | (0.45, -0.93)  | A         | No DIF        |
| qs34 | 1.48     | .22     | -0.43                       | (0.29, -1.15)  | A         | No DIF        |
| qs35 | 0.24     | .62     | -0.18                       | (0.54, -0.90)  | A         | No DIF        |
| qs36 | 3.70     | .05     | -0.72                       | (0.01, -1.45)  | A         | No DIF        |
| qs37 | 0.34     | .56     | -0.21                       | (0.50, -0.93)  | A         | No DIF        |
| qs38 | 0.24     | .62     | -0.18                       | (0.55, -0.91)  | A         | No DIF        |
| qs39 | 0.62     | .43     | 0.28                        | (0.97, -0.42)  | A         | No DIF        |
| qs40 | 0.44     | .51     | -0.22                       | (0.44, -0.87)  | A         | No DIF        |
| qs41 | 0.45     | .50     | 0.23                        | (0.91, -0.45)  | A         | No DIF        |
| qs42 | 1.20     | .27     | -0.40                       | (0.31, -1.11)  | A         | No DIF        |
| qs43 | 1.33     | .25     | -0.52                       | (0.36, -1.39)  | A         | No DIF        |
| qs44 | 2.29     | .13     | 0.61                        | (1.39, -0.18)  | A         | No DIF        |
| qs45 | 8.64     | .00     | 1.85                        | (3.08, 0.62)   | B+        | Favours Males |
| qs46 | 0.49     | .48     | 0.26                        | (0.98, -0.47)  | A         | No DIF        |
| qs47 | 1.82     | .18     | -0.53                       | (0.25, -1.30)  | A         | No DIF        |
| qs48 | 0.17     | .68     | 0.14                        | (0.82, -0.54)  | A         | No DIF        |

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta MH$ ) | 95% C.I.       | DIF Class | Direction       |
|------|----------|---------|-----------------------------|----------------|-----------|-----------------|
| qs49 | 15.76    | .00     | -1.52                       | (-0.76, -2.29) | B-        | Favours Females |
| qs50 | 0.10     | .75     | -0.12                       | (0.60, -0.84)  | A         | No DIF          |
| qs51 | 23.98    | .00     | -1.68                       | (-1.00, -2.35) | C-        | Favours Females |
| qs52 | 1.83     | .18     | -0.44                       | (0.20, -1.07)  | A         | No DIF          |
| qs53 | 0.11     | .74     | 0.12                        | (0.83, -0.59)  | A         | No DIF          |
| qs54 | 0.15     | .70     | -0.14                       | (0.58, -0.86)  | A         | No DIF          |
| qs55 | 0.35     | .55     | 0.20                        | (0.86, -0.46)  | A         | No DIF          |
| qs56 | 0.02     | .89     | 0.05                        | (0.76, -0.66)  | A         | No DIF          |
| qs57 | 0.11     | .74     | -0.11                       | (0.54, -0.76)  | A         | No DIF          |
| qs58 | 4.45     | .03     | -0.79                       | (-0.05, -1.53) | A         | No DIF          |
| qs59 | 2.30     | .13     | -0.48                       | (0.14, -1.10)  | A         | No DIF          |
| qs60 | 6.51     | .01     | -0.93                       | (-0.20, -1.66) | A         | No DIF          |

**Note:**  $\Delta MH$  = Delta Mantel–Haenszel effect size. A positive  $\Delta MH$  value indicates DIF favouring the reference group (Males), while a negative  $\Delta MH$  value indicates DIF favouring the focal group (Females). Classification Criteria: Category A (Negligible DIF):  $|\Delta MH| < 1.0$  and  $p > .05 \rightarrow$  No DIF item retained, Category B (Moderate DIF):  $1.0 \leq |\Delta MH| < 1.5$  and  $p \leq .05 \rightarrow$  Slight to moderate DIF flagged for review, Category C (Large DIF):  $|\Delta MH| \geq 1.5$  and  $p \leq .05 \rightarrow$  Moderate to large DIF indicating substantial bias

Table 1 revealed that 9 out of 60 items (15%) exhibited significant differential functioning between male and female students. Five items showed bias favouring male students, with item qs24 demonstrating the largest effect (C+ classification, ES = 1.81). Four items favoured female students, with item qs51 showing the strongest bias toward females (C- classification, ES = -1.68). The majority of items (51 items, 85%) functioned

equivalently across gender groups. This finding suggests that while the test was generally fair across gender, certain items may have contained content, wording, or contexts that differentially advantaged one gender group over the other, even when overall ability was controlled.

## Research Question Two

How many items in the BECE 2022 Basic Science Multiple Choice Test function differentially based on location in Zone C of Benue State, Nigeria?

To answer this research question, Cochran-Mantel-Haenszel DIF in jmetrik was used.

**Table 2 Presents the Cochran-Mantel-Haenszel DIF analysis results for location groups (Urban as reference group, Rural as focal group) based.**

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta MH$ ) | 95% C.I.       | DIF Class | Direction     |
|------|----------|---------|-----------------------------|----------------|-----------|---------------|
| qs1  | 111.58   | .00     | -5.85                       | (-4.45, -7.25) | C-        | Favours Rural |
| qs2  | 3.30     | .07     | 1.52                        | (3.24, -0.19)  | A         | No DIF        |
| qs3  | 0.13     | .72     | 0.20                        | (1.31, -0.91)  | A         | No DIF        |
| qs4  | 0.20     | .65     | -0.21                       | (0.70, -1.12)  | A         | No DIF        |
| qs5  | 11.17    | .00     | -2.28                       | (-0.85, -3.70) | B-        | Favours Rural |
| qs6  | 37.50    | .00     | 2.65                        | (3.53, 1.77)   | C+        | Favours Urban |
| qs7  | 133.91   | .00     | 5.08                        | (6.13, 4.03)   | C+        | Favours Urban |
| qs8  | 15.57    | .00     | 1.66                        | (2.51, 0.81)   | B+        | Favours Urban |
| qs9  | 42.62    | .00     | 2.64                        | (3.47, 1.81)   | C+        | Favours Urban |
| qs10 | 1.79     | .18     | 0.46                        | (1.18, -0.26)  | A         | No DIF        |
| qs11 | 53.87    | .00     | 3.07                        | (3.96, 2.19)   | C+        | Favours Urban |
| qs12 | 13.42    | .00     | -1.76                       | (-0.77, -2.74) | B-        | Favours Rural |
| qs13 | 1.18     | .28     | -0.46                       | (0.39, -1.30)  | A         | No DIF        |
| qs14 | 2.77     | .10     | 0.50                        | (1.17, -0.18)  | A         | No DIF        |
| qs15 | 31.83    | .00     | 2.59                        | (3.53, 1.64)   | C+        | Favours Urban |
| qs16 | 33.87    | .00     | -2.50                       | (-1.61, -3.39) | C-        | Favours Rural |
| qs17 | 0.47     | .49     | -0.35                       | (0.69, -1.39)  | A         | No DIF        |
| qs18 | 15.26    | .00     | 1.66                        | (2.49, 0.82)   | B+        | Favours Urban |
| qs19 | 57.50    | .00     | 3.04                        | (3.87, 2.22)   | C+        | Favours Urban |
| qs20 | 93.53    | .00     | -3.76                       | (-2.92, -4.59) | C-        | Favours Rural |

|      |       |     |       |                |    |               |
|------|-------|-----|-------|----------------|----|---------------|
| qs21 | 12.79 | .00 | 1.66  | (2.61, 0.72)   | B+ | Favours Urban |
| qs22 | 35.47 | .00 | -2.21 | (-1.45, -2.98) | C- | Favours Rural |
| qs23 | 0.22  | .64 | 0.22  | (1.11, -0.68)  | A  | No DIF        |
| qs24 | 0.42  | .52 | -0.29 | (0.60, -1.18)  | A  | No DIF        |
| qs25 | 6.81  | .01 | -1.15 | (-0.27, -2.03) | B- | Favours Rural |
| qs26 | 31.49 | .00 | 2.18  | (2.99, 1.38)   | C+ | Favours Urban |

| Item | $\chi^2$ | p-value | Effect Size<br>( $\Delta MH$ ) | 95% C.I.       | DIF<br>Class | Direction     |
|------|----------|---------|--------------------------------|----------------|--------------|---------------|
| qs27 | 3.56     | .06     | 0.69                           | (1.45, -0.07)  | A            | No DIF        |
| qs28 | 11.32    | .00     | -1.47                          | (-0.61, -2.33) | B-           | Favours Rural |
| qs29 | 39.22    | .00     | 2.68                           | (3.55, 1.80)   | C+           | Favours Urban |
| qs30 | 26.83    | .00     | 2.19                           | (3.04, 1.33)   | C+           | Favours Urban |
| qs31 | 4.71     | .03     | 1.14                           | (2.16, 0.11)   | B+           | Favours Urban |
| qs32 | 77.46    | .00     | -3.69                          | (-2.79, -4.60) | C-           | Favours Rural |
| qs33 | 28.86    | .00     | -2.25                          | (-1.41, -3.08) | C-           | Favours Rural |
| qs34 | 7.43     | .01     | 1.21                           | (2.12, 0.31)   | B+           | Favours Urban |
| qs35 | 80.07    | .00     | -3.79                          | (-2.90, -4.68) | C-           | Favours Rural |
| qs36 | 61.84    | .00     | -3.41                          | (-2.55, -4.27) | C-           | Favours Rural |
| qs37 | 28.35    | .00     | -2.33                          | (-1.44, -3.23) | C-           | Favours Rural |
| qs38 | 25.74    | .00     | -2.26                          | (-1.35, -3.17) | C-           | Favours Rural |
| qs39 | 73.82    | .00     | -3.28                          | (-2.46, -4.10) | C-           | Favours Rural |
| qs40 | 94.49    | .00     | -4.99                          | (-3.79, -6.18) | C-           | Favours Rural |
| qs41 | 6.75     | .01     | -1.15                          | (-0.28, -2.03) | B-           | Favours Rural |
| qs42 | 1.61     | .20     | 0.52                           | (1.38, -0.35)  | A            | No DIF        |
| qs43 | 0.60     | .44     | 0.33                           | (1.23, -0.57)  | A            | No DIF        |
| qs44 | 9.62     | .00     | 1.33                           | (2.23, 0.44)   | B+           | Favours Urban |
| qs45 | 0.02     | .88     | -0.11                          | (1.27, -1.48)  | A            | No DIF        |
| qs46 | 16.19    | .00     | -1.70                          | (-0.85, -2.56) | B-           | Favours Rural |
| qs47 | 69.12    | .00     | 3.70                           | (4.66, 2.75)   | C+           | Favours Urban |
| qs48 | 6.65     | .01     | 1.04                           | (1.85, 0.22)   | B+           | Favours Urban |
| qs49 | 0.38     | .54     | 0.26                           | (1.10, -0.59)  | A            | No DIF        |
| qs50 | 67.08    | .00     | 4.05                           | (5.10, 2.99)   | C+           | Favours Urban |
| qs51 | 28.47    | .00     | -2.18                          | (-1.36, -3.00) | C-           | Favours Rural |
| qs52 | 35.39    | .00     | 2.57                           | (3.43, 1.70)   | C+           | Favours Urban |
| qs53 | 2.75     | .10     | -0.73                          | (0.13, -1.59)  | A            | No DIF        |
| qs54 | 20.08    | .00     | -2.05                          | (-1.16, -2.95) | C-           | Favours Rural |
| qs55 | 6.95     | .01     | 1.12                           | (1.96, 0.27)   | B+           | Favours Urban |
| qs56 | 48.95    | .00     | -3.05                          | (-2.15, -3.95) | C-           | Favours Rural |
| qs57 | 0.18     | .67     | 0.18                           | (0.99, -0.63)  | A            | No DIF        |
| qs58 | 29.67    | .00     | 2.00                           | (2.80, 1.21)   | C+           | Favours Urban |
| qs59 | 24.15    | .00     | 1.80                           | (2.54, 1.05)   | C+           | Favours Urban |
| qs60 | 0.54     | .46     | 0.32                           | (1.18, -0.54)  | A            | No DIF        |

**Note:**  $\Delta MH$  = Delta Mantel–Haenszel effect size. A positive  $\Delta MH$  value indicates DIF favouring the reference group (Urban schools), while a negative  $\Delta MH$  value indicates DIF favouring the focal group (Rural schools). Classification Criteria: Category A (Negligible DIF):  $|\Delta MH| < 1.0$  and  $p > .05 \rightarrow$  No DIF item retained Category B (Moderate DIF):  $1.0 \leq |\Delta MH| < 1.5$  and  $p \leq .05 \rightarrow$  Slight to moderate DIF flagged for review Category C (Large DIF):  $|\Delta MH| \geq 1.5$  and  $p \leq .05 \rightarrow$  Moderate to large DIF indicating substantial bias

Table 2 revealed a substantial level of differential item functioning based on school location. A striking 42 out of 60 items (70%) exhibited significant DIF between urban and rural students. Of these, 25 items favoured rural students while 17 items favoured urban students. Thirty items showed moderate to large DIF (Category C), indicating substantial bias. The magnitude of some effect sizes was extremely large, such as item qs1 (ES = -5.85 favouring rural), item qs7 (ES = 5.08 favouring urban), and item qs40 (ES = -4.99 favouring rural). This finding suggests serious equity concerns in the test, as a majority of items functioned differently for students from urban versus rural educational settings. This bias could reflect differences in curriculum exposure, teaching quality, resource availability, or cultural contexts between urban and rural schools that may advantage one group over the other beyond their actual Basic Science proficiency.

### **Research Question Three**

How many items in the BECE 2022 Basic Science Multiple Choice Test function differentially based on school types (co-educational and non-co-educational education) in Zone C of Benue State, Nigeria?

To answer this research question, Cochran-Mantel-Haenszel DIF in jmetrik was used.

**Table 3 Presents the Cochran-Mantel-Haenszel DIF analysis results for school type groups (Co-educational as reference group, Non-co-educational education as focal group).**

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta$ MH) | 95% C.I.       | DIF Class | Direction                  |
|------|----------|---------|----------------------------|----------------|-----------|----------------------------|
| qs1  | 13.61    | .00     | 6.28                       | (10.87, 1.70)  | C+        | Favours Co-ed              |
| qs2  | 0.02     | .88     | 0.27                       | (3.81, -3.27)  | A         | No DIF                     |
| qs3  | 1.59     | .21     | -1.21                      | (0.70, -3.11)  | A         | No DIF                     |
| qs4  | 0.04     | .84     | -0.15                      | (1.27, -1.56)  | A         | No DIF                     |
| qs5  | 2.28     | .13     | -1.34                      | (0.47, -3.15)  | A         | No DIF                     |
| qs6  | 17.20    | .00     | 2.19                       | (3.27, 1.11)   | C+        | Favours Co-ed              |
| qs7  | 15.41    | .00     | -4.17                      | (-1.82, -6.53) | C-        | Favours Non-co-educational |
| qs8  | 0.00     | 1.00    | -0.00                      | (1.26, -1.27)  | A         | No DIF                     |
| qs9  | 14.57    | .00     | -1.92                      | (-0.87, -2.96) | B-        | Favours Non-co-educational |
| qs10 | 0.06     | .81     | -0.11                      | (0.83, -1.06)  | A         | No DIF                     |
| qs11 | 33.65    | .00     | -2.94                      | (-1.88, -4.00) | C-        | Favours Non-co-educational |
| qs12 | 0.81     | .37     | 0.62                       | (2.00, -0.76)  | A         | No DIF                     |
| qs13 | 4.36     | .04     | 1.21                       | (2.36, 0.05)   | B+        | Favours Co-ed              |
| qs14 | 0.44     | .51     | -0.35                      | (0.73, -1.44)  | A         | No DIF                     |
| qs15 | 9.01     | .00     | 1.82                       | (3.06, 0.58)   | B+        | Favours Co-ed              |
| qs16 | 43.88    | .00     | 3.52                       | (4.63, 2.42)   | C+        | Favours Co-ed              |
| qs17 | 13.40    | .00     | -2.23                      | (-0.96, -3.50) | B-        | Favours Non-co-educational |
| qs18 | 2.91     | .09     | 0.87                       | (1.89, -0.14)  | A         | No DIF                     |
| qs19 | 139.44   | .00     | -7.07                      | (-5.56, -8.58) | C-        | Favours Non-co-educational |
| qs20 | 0.00     | .99     | -0.00                      | (1.07, -1.08)  | A         | No DIF                     |
| qs21 | 28.25    | .00     | -2.71                      | (-1.58, -3.84) | C-        | Favours Non-co-educational |
| qs22 | 94.07    | .00     | 5.57                       | (6.87, 4.26)   | C+        | Favours Co-ed              |

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta$ MH) | 95% C.I.       | DIF Class | Direction                  |
|------|----------|---------|----------------------------|----------------|-----------|----------------------------|
| qs23 | 15.74    | .00     | 2.01                       | (3.02, 1.00)   | B+        | Favours Co-ed              |
| qs24 | 111.24   | .00     | 5.22                       | (6.28, 4.15)   | C+        | Favours Co-ed              |
| qs25 | 1.61     | .20     | 1.06                       | (2.72, -0.60)  | A         | No DIF                     |
| qs26 | 21.74    | .00     | -2.63                      | (-1.47, -3.79) | C-        | Favours Non-co-educational |
| qs27 | 0.03     | .85     | -0.11                      | (1.00, -1.21)  | A         | No DIF                     |
| qs28 | 3.20     | .07     | 1.13                       | (2.37, -0.11)  | A         | No DIF                     |
| qs29 | 105.05   | .00     | -5.32                      | (-4.14, -6.50) | C-        | Favours Non-co-educational |
| qs30 | 104.76   | .00     | 4.67                       | (5.71, 3.62)   | C+        | Favours Co-ed              |
| qs31 | 0.96     | .33     | 0.72                       | (2.19, -0.74)  | A         | No DIF                     |
| qs32 | 40.12    | .00     | 3.42                       | (4.54, 2.29)   | C+        | Favours Co-ed              |
| qs33 | 0.04     | .83     | 0.12                       | (1.29, -1.05)  | A         | No DIF                     |
| qs34 | 14.15    | .00     | -2.93                      | (-1.33, -4.53) | C-        | Favours Non-co-educational |
| qs35 | 2.59     | .11     | 1.12                       | (2.51, -0.27)  | A         | No DIF                     |
| qs36 | 0.37     | .54     | -0.43                      | (0.97, -1.82)  | A         | No DIF                     |
| qs37 | 3.43     | .06     | 1.20                       | (2.47, -0.07)  | A         | No DIF                     |
| qs38 | 18.34    | .00     | 3.25                       | (4.76, 1.74)   | C+        | Favours Co-ed              |
| qs39 | 4.49     | .03     | 1.24                       | (2.44, 0.04)   | B+        | Favours Co-ed              |
| qs40 | 54.99    | .00     | -3.78                      | (-2.60, -4.95) | C-        | Favours Non-co-educational |
| qs41 | 23.11    | .00     | 2.37                       | (3.37, 1.37)   | C+        | Favours Co-ed              |
| qs42 | 40.81    | .00     | -4.49                      | (-2.94, -6.03) | C-        | Favours Non-co-educational |
| qs43 | 8.68     | .00     | -2.24                      | (-0.68, -3.80) | B-        | Favours Non-co-educational |
| qs44 | 59.04    | .00     | 4.11                       | (5.22, 3.00)   | C+        | Favours Co-ed              |
| qs45 | 25.31    | .00     | 3.87                       | (5.59, 2.15)   | C+        | Favours Co-ed              |
| qs46 | 1.29     | .26     | 0.77                       | (2.12, -0.57)  | A         | No DIF                     |
| qs47 | 4.36     | .04     | -1.99                      | (-0.03, -3.95) | B-        | Favours Non-               |

| Item | $\chi^2$ | p-value | Effect Size ( $\Delta MH$ ) | 95% C.I.        | DIF Class | Direction                  |
|------|----------|---------|-----------------------------|-----------------|-----------|----------------------------|
| qs48 | 2.89     | .09     | 1.37                        | (2.96, -0.22)   | A         | co-educational<br>No DIF   |
| qs49 | 12.80    | .00     | -2.08                       | (-0.89, -3.28)  | B-        | Favours Non-co-educational |
| qs50 | 3.34     | .07     | -1.16                       | (0.13, -2.44)   | A         | No DIF                     |
| qs51 | 197.33   | .00     | -8.71                       | (-6.75, -10.68) | C-        | Favours Non-co-educational |
| qs52 | 0.17     | .68     | 0.19                        | (1.13, -0.75)   | A         | No DIF                     |
| qs53 | 6.11     | .01     | -1.62                       | (-0.28, -2.95)  | B-        | Favours Non-co-educational |
| qs54 | 1.39     | .24     | 0.66                        | (1.81, -0.49)   | A         | No DIF                     |
| qs55 | 42.25    | .00     | 4.43                        | (5.94, 2.92)    | C+        | Favours Co-ed              |
| qs56 | 7.24     | .01     | -1.70                       | (-0.39, -3.01)  | B-        | Favours Non-co-educational |
| qs57 | 10.72    | .00     | 1.65                        | (2.70, 0.59)    | B+        | Favours Co-ed              |
| qs58 | 2.38     | .12     | -1.47                       | (0.48, -3.42)   | A         | No DIF                     |
| qs59 | 23.13    | .00     | -2.67                       | (-1.52, -3.81)  | C-        | Favours Non-co-educational |
| qs60 | 0.32     | .57     | 0.38                        | (1.74, -0.97)   | A         | No DIF                     |

**Note:**  $\Delta MH$  = Delta Mantel–Haenszel effect size. A positive  $\Delta MH$  value indicates DIF favouring the reference group (Co-educational schools), while a negative  $\Delta MH$  value indicates DIF favouring the focal group (Non-co-educational schools). Classification Criteria: Category A (Negligible DIF):  $|\Delta MH| < 1.0$  and  $p > .05 \rightarrow$  No DIF item retained Category B (Moderate DIF):  $1.0 \leq |\Delta MH| < 1.5$  and  $p \leq .05 \rightarrow$  Slight to moderate DIF flagged for Category C (Large DIF):  $|\Delta MH| \geq 1.5$  and  $p \leq .05 \rightarrow$  Moderate to large DIF indicating substantial bias

Table 3 revealed that 33 out of 60 items (55%) demonstrated significant differential functioning between co-educational and non-co-educational schools. Eighteen items favoured students

from co-educational schools while 15 items favoured students from non-co-educational schools. Twenty-one items exhibited moderate to large DIF (Category C), with some showing extremely large effect sizes such as item qs51 (ES = -8.71 favouring co-educational), item qs19 (ES = -7.07 favouring co-educational), item qs1 (ES = 6.28 favouring non-co-educational), and item qs22 (ES = 5.57 favouring non-co-educational). The substantial proportion of items showing DIF and the magnitude of bias observed indicate that school type represents a significant source of measurement inequity in this test. These differences may reflect variations in instructional approaches, learning environments, peer dynamics, or curriculum emphasis between co-educational and non-co-educational educational settings that influence how students respond to certain items independent of their true Basic Science ability.

## **Discussion of Findings**

The CMH analysis revealed that 9 of 60 items (15%) exhibited statistically significant gender-based DIF, with five items (qs1, qs16, qs24, qs30, and qs45) favouring male students and four (qs17, qs19, qs49, and qs51) favouring females. This rate falls within the moderate range documented in the Nigerian psychometric literature. Earlier studies reported considerably higher 28% and 20% for Mathematics and English Language respectively (Ajeigbe & Afolabi, 2014), and 34% for NABTEB Mathematics items (Omorogiuwa & Iro-Aghedo, 2016) while more recent studies have converged toward figures comparable to the present finding, ranging from 6% to 16% (Effiom, 2021; Obiageli & Ebouh, 2022; Olutola & Ihechu, 2023). The near-balanced distribution of DIF items across gender groups suggests the absence of systematic bias, which is encouraging from a fairness standpoint. However, two items classified as Category C qs24 (ES = 1.81, favouring males) and qs51 (ES = -1.68, favouring females) warrant concern, as their effect sizes are large enough to meaningfully distort individual scores. These items require content review to identify linguistic, contextual, or

cognitive features that may differentially advantage one gender group independent of overall ability.

Location-based DIF emerged as the most severe source of measurement inequity in this study. Forty-two of 60 items (70%) functioned differentially between urban and rural students 25 favouring rural students and 17 favouring urban students with 30 items classified as Category C. Several items recorded effect sizes of considerable magnitude, notably qs1 ( $ES = -5.85$ , favouring rural), qs7 ( $ES = 5.08$ , favouring urban), and qs40 ( $ES = -4.99$ , favouring rural). This rate substantially exceeds figures from comparable studies, which documented location-based DIF ranging from 32% to 58% (Olutola et al., 2020; Evans et al., 2022; Obiageli & Ebouh, 2022), indicating an unusually severe level of geographical measurement inequity within Zone C.

These findings likely reflect persistent structural disparities between urban and rural schools in Nigeria including differential access to laboratory facilities, instructional materials, qualified teachers, and curriculum implementation quality, which produce unequal exposure to the scientific concepts assessed by the test. The bidirectional pattern of DIF, with items favouring both locations, further suggests that different items tap into forms of experiential knowledge that are unequally distributed across geographical settings rather than reflecting a common underlying proficiency. The extremely large effect sizes recorded for several items indicate that these items function almost as parallel instruments for urban and rural examinees, fundamentally compromising score comparability across locations.

Thirty-three of 60 items (55%) functioned differentially between students from co-educational and non-co-educational schools, with 18 items favouring co-educational students and 15 favouring those from non-co-educational institutions. Twenty-one items were classified as Category C, with several displaying notably large effect sizes: qs51 ( $ES = -8.71$ ), qs19 ( $ES = -7.07$ ), qs1 ( $ES$

= 6.28), and qs22 (ES = 5.57). School type thus emerges as a substantial and underexplored source of measurement inequity in the Nigerian basic education context.

Unlike gender and location DIF, school type DIF has received limited empirical attention in the literature, making these findings particularly significant. The fact that over half of all items functioned differentially suggests that differences in pedagogical approaches, classroom dynamics, institutional culture, and student socialization between co-educational and non-co-educational schools meaningfully shape item response behaviour beyond true ability differences. The bidirectional bias pattern indicates that neither school type is uniformly advantaged; rather, different items align differently with the instructional experiences characteristic of each environment. The extremely large effect sizes associated with several items raise the possibility that these items measure qualitatively different constructs across school types, calling into question the comparability of BECE Basic Science scores across school environments and highlighting the need for school-type-sensitive fairness considerations in future test development.

## **Conclusion**

Based on the findings of this study, the following conclusions were drawn.

The BECE 2022 Basic Science and Technology Achievement Test administered to Upper Basic III students in Zone C of Benue State exhibited varying degrees of differential item functioning across the three demographic subgroups examined. Gender-based DIF, while present in 15% of items, was comparatively modest and showed no systematic directional bias, suggesting that the test is reasonably though not entirely fair across gender groups. However, the presence of two Category C items indicates that residual gender-related bias exists and requires attention during item review. Location-based DIF was the most critical threat to test fairness identified in this study. With 70% of items

functioning differentially between urban and rural students and numerous items carrying large effect sizes, the test cannot be considered a fair or comparable measure of Basic Science proficiency across geographical settings in Zone C. Score differences between urban and rural students are likely to reflect structural inequalities in educational opportunity rather than true differences in ability. School type also constituted a substantial source of measurement inequity, with 55% of items exhibiting DIF between co-educational and non-co-educational school students. The magnitude of bias observed including several items with extremely large effect sizes raises serious concerns about the comparability of scores across school environments and suggests that institutional context meaningfully shapes how students engage with test items beyond their actual proficiency in Basic Science.

### **Recommendations**

Based on the findings of this study, the following recommendations are proposed for various stakeholders in the educational assessment system.

1. The Benue State Examinations Board (BSEB) should institutionalize DIF analysis as a mandatory component of its post-examination review process. Items classified as Category C should be flagged for immediate content review or removal from future item banks before they influence high-stakes decisions on student placement, certification, and progression.
2. The Federal and Benue State Ministries of Education should address structural inequalities in educational provision between urban and rural schools through targeted investment in laboratory facilities, instructional materials, and qualified science teachers in underserved areas. Equitable curriculum implementation across all school types and locations is essential for ensuring that test scores reflect genuine ability rather than differences in educational opportunity.

## References

- Adejoh & Obinne (2015). *Basic Issues in Tests, Measurement and Evaluation (1<sup>st</sup> Edition)*. Makurdi: Asofad Publishers.
- Ajeigbe, T. O., & Afolabi, E. R. I. (2014). Assessing unidimensionality and differential item functioning in qualifying examination for senior secondary school students in Osun State, Nigeria. *Journal of Educational Research*, 4(4), 30–37.
- Chalmers, R. P. (2018). Model-based measures for detecting and quantifying response bias. *Psychometrika*, 83(3), 696–732.
- Effiom, A. P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River State, Nigeria. *Global Journal of Educational Research*, 20(1), 55–62.
- Emaikwu, S. O. (2019). *Fundamentals of research methodology and statistics* (Rev. ed.). Selfers Academic Press.
- Evans, G. U., Uko, M. P., & Ekim, R. E. D. (2022). Investigation of differential item functioning of Basic Education Certificate Examination (BECE) mathematics items in Akwa Ibom State, Nigeria. *African Journal of Behavioural and Scale Development Research*, 4(1). (Complete page numbers required.)
- Ihechu, K.J.P. & Madu, A.O., (2023) Differential Item Functioning of National Examinations Council (NECO) Agricultural Science Questions For (2015). *Journal of Educational and Leadership*
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266–283.
- Madueke, U. O., & Ebouh, C. N. (2022). Differential item functioning of WAEC Senior Secondary Certificate Examination Biology multiple choice items. *British International Journal of Education and Social Sciences*, 9(6).

- Magis, D., Béland, S., & Raîche, G. (2020). Detection of differential item functioning using R. In W. J. van der Linden (Ed.), *Handbook of item response theory* (Vol. 2, pp. 405–425). Chapman & Hall/CRC.
- Martinková, P., & Drabinová, A. (2018). Shiny Item Analysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal*, 10(2), 503–515.
- Nigerian Educational Research and Development Council. (2016). *Basic science and technology curriculum for Upper Basic Education*. NERDC Press.
- Olutola, A.T. (2020). School Location and Gender as Predictors of Students' Performance in WASSCE Multiple Choice Test in Biology. *Liceo Journal of Higher Education Research*, 12 (1), 1-16.
- Omorogiwa, K. O., & Iro-Aghedo, E. P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examinations Board (NABTEB) 2015 mathematics multiple choice examination. *International Journal of Education, Learning and Development*, 4(10), 25–35.
- Osterlind, S. J., & Everson, H. T. (2019). *Differential item functioning in Educational measurement* (5th ed.). Routledge.
- Owolabi, H.O. (2019). Test, Measurement and Evaluation. Introduction to Educational Measurement and Evaluation. Test, Measurement and Evaluation.(PP.93-99). Department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin.
- Popham, W. J. (2016). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.
- Oguguo, B. C. E., & Lotobi, R. A. (2019). Parameters of Basic Science test items of 2011 Basic Education Certificate Examination using Item Response Theory (IRT) approach in Delta State, Nigeria. *European Journal of Educational Sciences*, 6(1), 22–36.